Diese Website verwendet Cookies von Google, um Dienste anzubieten und den Traffic zu analysieren. Deine IP-Adresse und dein User-Agent werden zusammen mit Messwerten zur Leistung und Sicherheit für Google freigegeben. So können Nutzungsstatistiken generiert, Missbrauchsfälle erkannt und behoben und die Qualität des Dienstes gewährleistet werden.

WEITERE INFORMATIONEN OK

Tuesday, February 06, 2007

Falsifiable design: a methodology for evaluating theoretical technologies

Theoretical or futuristic technologies have long been a staple of science fiction. Increasingly the future of technology development has become an important political issue as decisions are made to fund a variety of alternative energy and emission reduction technologies in response to projected long-term climate changes, medical technologies to address health problems and epidemic threats, military R&D in response to recent and feared future security threats, and so on. Hundreds of billions of dollars of R&D funding over the next decade hinges on questions of future technologies. Some of these technologies in some of their forms may be harmful or dangerous; can we evaluate these dangers before spending billions of dollars of development and capital investment in technologies that may end up expensively restricted or banned as designed?

We lack a good discipline of theoretical technology. As a result, discussion of such technologies among scientists and engineers and laypeople alike often never gets beyond naive speculation, which ranges from dismissing or being antagonistic to such possibilities altogether to proclamations that sketchy designs in hard-to-predict areas of technology are easy and/or inevitable and thus worthy of massive attention and R&D funding. At one extreme are recent news stories, supposedly based on the new IPCC report on global warming, that claim that we will not be able to undo the effects of global warming for at least a thousand years. At the other extreme are people like Eric Drexler and Ray Kurzweil who predict runaway technological progress within our lifetimes, allowing us to not only easily and quickly reverse global warming but also conquer death, colonize space, and much else. Such astronomically divergent views, each held by highly intelligent and widely esteemed scientists and engineers, reflect vast disagreements about the capabilities of future technology. Today we lack good ways for proponents of such views to communicate in a disciplined way with each other or to the rest of us about such claims and we lack good ways as individuals, organizations, and as a society to evaluate them.

Sometimes descriptions of theoretical technologies lead to popular futuristic movements such as extropianism, transhumanism, cryonics, and many others. These movements make claims, often outlandish but not refuted by scientific laws, that the nature of technologies in the future should influence how we behave today. In some cases this leads to quite dramatic changes in outlook and behavior. For example, people ranging from baseball legend Ted Williams to cryptography pioneer and nanotechnologist Ralph Merkle have spent up to a hundred thousand dollars or more to have their bodies frozen in the hopes that future technologies will be able to fix the freezing damage and cure whatever killed them. Such beliefs can radically alter a person's outlook on life, but how can we rationally evaluate their credibility?

Some scientific fields have developed that inherently involve theoretical technology. For example, SETI (and less obviously its cousin astrobiology) inherently involve speculation about what technologies hypothetical aliens might possess.

About Me

Nick Szabo

"A premier thinker about history, law and economics, and the lessons they have for security." -- Adam Shostack, Emergent Chaos

"Szabo comes out with these essays that leave me in awe." -- Brian Dunbar

"Reading material that is eclectic, challenging and endlessly fascinating." -- Sean McGrath, Propylon

"Like most blogs worth my attention, this blog is updated only infrequently. That is because the authors of blogs worth my attention only post when they have something to say that is true, relevant and not already known by their audience. Most of the human race does not have the skill to know when an idea has these three properties. The skill is particularly rare in the fields of politics and economics, which is why this blog is such a rare and valuable thing." -- Richard Hollerith

View my complete profile

Eric Drexler called the study of such theoretical technologies "exploratory engineering" or "theoretical applied science." Currently futurists tend to evaluate such designs based primarily on their mathematical consistency with known physical law. But mere mathematical consistency with high-level abstractions is woefully insufficient, except in computer science or where only the simplest of physics is involved, for demonstrating the future workability or desirability of possible technologies. And it falls far short of the criteria engineers and scientists generally use to evaluate each others' work.

Traditional science and engineering methods are ill-equipped to deal with such proposals. As "critic" has pointed out, both demand near-term experimentation or tests to verify or falsify claims about the truth of a theory or whether a technology actually works. Until such an experiment has occurred a scientific theory is a mere hypothesis, and until a working physical model, or at least a description sufficient to create such has been produced, a proposed technology is merely theoretical. Scientists tend to ignore theories that they have no prospect of testing and engineers tend to ignore designs that have no imminent prospect of being built, and in the normal practice of these fields this is a quite healthy attitude. But to properly evaluate theoretical technologies we need ways other than near-term working models to reduce the large uncertainties in such proposals and to come to some reasonable estimates of their worthiness and relevance to R&D and other decisions we make today.

The result of these divergent approaches is that when scientists and engineers talk to exploratory engineers they talk past each other. In fact, neither side has been approaching theoretical technology in a way that allows it to be properly evaluated. Here is one of the more productive such conversations -- the typical one is even worse. The theoreticians' claims are too often untestable, and the scientists and engineers too often demand inappropriately that descriptions be provided that would allow one to actually build the devices with today's technology.

To think about how we might evaluate theoretical technologies, we can start by looking a highly evolved system that technologists have long used to judge the worthiness of new technological ideas -- albeit not without controversy -- the patent system.

The patent system sets up several basic requirements for proving that an invention is worthy enough to become a property right: novelty, non-obviousness, utility, and enablement. (Enablement is also often called "written description" or "sufficiency of disclosure", although sometimes these are treated as distinct requirements that we need not get into for our purposes). Novelty and non-obviousness are used to judge whether a technology would have been invented anyway and are largely irrelevant for our purposes here (which is good news because non-obviousness is the source of most of the controversy about patent law). To be worthy of discussion in the context of, for example, R&D funding decisions, the utility of a theoretical technology, if or when in the future it is made to work, should be much larger than that required for a patent -- the technology if it works should be at least indirectly of substantial expected economic importance. I don't expect meeting this requirement to be much of a problem, as futurists usually don't waste much time on the economically trivial, and it may only be needed to distinguish important theoretical technologies from those merely found in science fiction for the purposes of entertainment.

The patent requirement that presents the biggest problem for theoretical technologies is *enablement*. The inventor or patent agent must write a description of the technology detailed enough to allow another engineer in the field to, given a perhaps large but not infinite amount of resources, build the invention and make it work. If the design is not mature enough to build -- if, for example, it requires further novel and non-obvious inventions to build it and make it work -- the law deems it of too little importance to entitle the proposer to a property right in it. Implementing the description must not require so much experimentation that it becomes a research project in its own right.

An interesting feature of enablement law is the distinction it makes between fields that are considered more predictable or less so. Software and (macro-)mechanical patents generally fall under the "predictable arts" and require must less detailed description to be enabling. But biology, chemistry, and many other "squishier" areas are much less predictable and thus require far more detailed description to be considered enabling.

Blog Archive

- ► 2018 (1)
- ► 2017 (3)
- ► 2016 (4)
- ► 2015 (3)
- ► 2014 (3)
- ► 2013 (3)
- ► 2012 (8)
- ► 2011 (12)
- ► 2010 (9)
- ► 2009 (29)
- ► 2008 (55)
- ▼ 2007 (47)
  - ► December (1)
  - ► October (2)
  - ► September (8)
  - ► August (10)
  - ► July (4)
  - ► June (4)
  - ► May (8)
  - ► April (1)
  - ► March (2)
  - ▼ February (4)
    - Partially self-replicating machines being built
    - Falsifiable design: a methodology for evaluating t...
    - Mining the vasty deep (ii)
    - Mining the vasty deep (i)
  - ► January (3)
- ► 2006 (130)
- ► 2005 (44)

Theoretical technology lacks enablement. For example, no astronautical engineer today can take Freeman Dyson's description of a Dyson Sphere and go out and build one, even given a generous budget. No one has been able to take Eric Drexler's *Nanosystems* and go out and build a molecular assembler. From the point of view of the patent system and of typical engineering practice, these proposals either require far more investment than is practical, or far have too many engineering uncertainties and dependencies on other unavailable technology to build them, or both, and thus are not worth considering.

For almost all scientists and engineers, the lack of imminent testability or enablement puts such ideas out of their scope of attention. For the normal practice of science and engineering this is healthy, but it can have quite unfortunate consequences when decisions about what kinds of R&D to fund are made or about how to respond to long-term problems. As a society we must deal with long-term problems such as undesired climate change. To properly deal with such problems we need ways to reason about what technologies may be built to address them over the spans of decades, centuries, or even longer over which we reason about and expect to deal with such problems. For example, the recent IPCC report on global warming has projected trends centuries in the future, but has not dealt in any serious or knowledgeable way with how future technologies might either exacerbate or ameliorate such trends over such a long span of time -- even though the trends they discuss are primarily caused by technologies in the first place.

Can we relax the enablement requirement to create a new system for evaluating theoretical technologies? We might call the result a "non-proximately enabled invention" or a "theoretical patent." [UPDATE: note that this is not a proposal for an actual legally enforceable long-term patent. This is a thought experiment for the purposes of seeing to what extent patent-like criteria can be used for evaluating theoretical technology]. It retains (and even strengthens) the utility requirement of a patent, but it relaxes the enablement requirement in a disciplined way. Instead of enabling a fellow engineer to build the device, the description must be sufficient to enable an research project -- a series of experiments and tests that would if successful lead to a working invention, and if unsuccessful would serve to demonstrate that the technology as proposed probably can't be made to work.

It won't always be possible to test claims about theoretical designs. They may be of a nature as to require technological breakthroughs to even test their claims. In many cases this will be due the lack of appreciation on the part of the theoretical designer of the requirement that such designs be testable. In some cases it will be the lack of skill or imagination of the theoretical designer to fail to be able to come up with such tests. In some cases it will be because the theoretical designer refuses to admit the areas of uncertainty in the design and thus denies the need for experiment to reduce these uncertainties. In some cases it may just be inherent in the radically advanced nature of the technology being studied that there is no possible way to test it without making other technology advances. Regardless of the reason the designs are untestable, these designs cannot be considered to be anything but entertaining science fiction. In only the most certain of areas -- mostly just in computer science and in the simplest and most verified of physics -- should mathematically valid but untested claims be deemed credible for the purposes of making important decisions.

Both successful and unsuccessful experiments that reduce uncertainty are valuable. Demonstrating that the technology can work as proposed allows lower-risk investments in it to be made sooner, and leads to decisions better informed about the capabilities of future technology. Demonstrating that it can't work as proposed prevents a large amount of money from being wasted trying to build it and prevents decisions from being made today based on the assumption that it will work in the future.

The distinction made by patent systems between fields that are more versus less predictable becomes even more important for theoretical designs. In computer science the gold standard is mathematical proof rather than experiment, and given such proof no program of experiment need be specified. Where simple physics is involved, such as the orbital mechanics used to plan space flights, the uncertainty is also usually very small and proof based on known physical laws is generally sufficient to show the future feasibility of orbital trajectories. (Whether the engines can be built to achieve such trajectories is another matter). Where simple scaling processes are involved (e.g. scaling up rockets from Redstone sized to Saturn sized) uncertainties are relatively small. Thus, to take our space flight example, the Apollo program had a relatively low uncertainty well ahead of time, but it was unusual in this regard. As soon as we get into dependency on new materials (e.g. the infamous "unobtainium" material stronger than any known materials), new chemical reactions or new ways of performing chemical reactions, and so on, things get murkier far more quickly, and it is essential to good

theoretical design that these uncertainties be identified and that experiments to reduce them be designed.

In other words, the designers of a theoretical technology in any but the most predictable of areas should identify its assumptions and claims that have not already been tested in a laboratory. They should design not only the technology but also a map of the uncertainties and edge cases in the design and a series of such experiments and tests that would progressively reduce these uncertainties. A proposal that lacks this admission of uncertainties coupled with designs of experiments that will reduce such uncertainties should not be deemed credible for the purposes of any important decision. We might call this requirement a requirement for a *falsifiable design*.

Falsifiable design resembles the systems engineering done with large novel technology programs, especially those that require large amounts of investment. Before making the large investments tests are conducted so that the program, if it won't actually work, will "fail early" with minimal loss, or will proceed with components and interfaces that actually work. Theoretical design must work on the same principle, but on longer time scales and with even greater uncertainty. The greater uncertainties involved make it even more imperative that uncertainties be resolved by experiment.

The distinction between a complete blueprint or enabling description and a description for the purposes of enabling falsifiability of a theoretical design is crucial. A patent enablement or an architectural blueprint is an a description of *how to build.* An enablement of falsifiability is a description of *if it were built -- and we don't claim to be describing or even to know how to build it -- there is how we believe it would function, here are the main uncertainties on how it would function, and here are the experiments that can be done to reduce these uncertainties.*

Good theoretical designers should be able to recognize the uncertainties in their own and others' designs. They should be able to characterize the uncertainties in such a way as to that evaluators to assign probabilities to the uncertainties and to decide what experiments or tests can most easily reduce the uncertainties. It is the recognition of uncertainties and the plan of uncertainty reduction, even more than the mathematics demonstrating the consistency of the design with physical law, that will enable the theoretical technology for the purposes, not of building it today, but of making important decisions today based on to what degree we can reasonably expect it to be built and produce value in the future.

Science fiction is an art form in which the goal is to suspend the reader's disbelief by artfully avoiding mention of the many uncertainties involved in the wonderful technologies. To be useful and credible -- to achieve truth rather than just the comfort or entertainment of suspension of disbelief -- theoretical design must do the opposite -- it must highlight the uncertainties and emphasize near-term experiments that can be done to reduce them.

Posted by Nick Szabo at 9:20 AM

10 comments:

albatross said...

This makes me think of some of Zubrin's discussion of the Mars Direct idea. He pointed out that the riskiest-looking part of the design (the ability to make fuel for your return voyage from nuclear power + atmosphere + feedstock brought along) was actually very easy to test.

The other interesting part of your discussion is that it lets us isolate the parts of some design we can't test, and make our informed

bets on those parts. And it seems like some good technologies will have the property that they mostly depend on stuff that can't be tested cheaply. (Think of stuff that scales in weird ways.)

11:42 AM

Nick Szabo said...

Speaking of betting, the well-defined claims of uncertain outcome that I call for here would also make good candidates for prediction markets, with the caveat of a couple limitations. First, people have more incentive to participate in prediction markets when the event(s) being bet on (here the experiment(s)) are imminent. Second, if the issue is sufficiently uncertain market participants will lack good information and the resulting prediction will merely reflect the general ignorance.

However, in the intermediate cases where this is good information but it is scattered, and for issues of sufficient economic importance or entertainment value prediction markets might work quite well.

I'd be interested in pointers to any such prediction markets that are currently active.

To the extent the uncertain parts of technologies can't be tested easily and imminently, they remain highly uncertain and make very poor bases for policy or other important decisions.

3:30 PM

A said...

as usual, i'm going to take exception to the idea. it strikes me as contrary to the fundamental notion that patents are intended to protect development, not encourage the protection of ideas per se. submarine patents are enough of a problem in the current system; why wouldn't this structure create a chilling effect on futuristic technologies? at the very least it strikes me as creating a fuzzy legal area on the bounds of theoretical technologies, which is exactly what you don't want in an area that's admittedly doing high-risk, low-return development (arguably also a big problem at the boundaries of software patent law).

so why do we need a patent protection program for the Neal Stephensons of the world? i'm not aware that they engage in a lot of empirical R&D, nor that they're being disincentivized at the moment (heck, Drexler even has his own Institute).

i think that to some extent you're drawing a false equivalence between renewable energy technologies (which are fairly well-bounded) and, say, eric drexler's nanobots -- i'm pretty sure we can agree that they're not talking about the same thing.

while the discussion is interesting on an academic level, the theory is rarely as important as the enabling technology: to take your somewhat far-out example, we can lay out many of the specifications of a Dyson sphere, and perhaps even many of the "theoretical uncertainties," but all of that does us little good until someone actually manages to build a Dyson sphere, and at that point they'll have de facto met their enablement requirement.

i'd like to think that this is an issue that transcends our politics. :) why not, for example, an incubator and/or VC fund specifically geared toward developing and actualizing theoretical technologies (that is, bridging the gap between the uncertainties you discuss and the actual enablement requirements)?

7:36 PM

Nick Szabo said...

Alas, I left a big misimpression. While establshing a new kind of legal protection for these ideas might be interesting in its own right -- but quite probably would work out badly for the reasons A points out -- that wasn't at all my motivation for suggesting this new kind of "patent". My motivation is simply to use patent-like criteria, but substantially altered as suggested, to evaluate theoretical technologies for the purposes of making policy decisions, e.g to figure out how to respond to global warming and to figure out what

kinds of science and R&D to fund.

The question is to what extent policy makers should take these futuristic proposals seriously -- for example should we take Drexler's scenario seriously when making decades- or even centuries-long projections or plans to respond to global warming? My short answer is that we should only take it seriously if but only if it is specified in an imminently falsifiable manner.

I'm not sure what you mean when you say that renewable energy technologies are "fairly well bounded," but it doesn't sound right. There are major uncertainties in many of them that reside in one of those quite unpredictable arts, chemistry, as does the core of Drexler's proposals. Indeed Drexler's proposal might be less uncertain because it involves mostly one big chemical uncertainty (the mechanosynthesis question) surrounded by quite a whole lot of two quite more predictable arts, mechanical and software engineering. It has a probably very big payoff (if it works) rather than a very large number of uncertainties with small payoffs. Granted, the Drexler uncertainty could easily end up being that it doesn't work at all while we already know many alternative technologies work somewhat well -- if that's all you meant by well-bounded then I agree.

VC's aren't going to invest in things that they can't imminently build -- you'd need to have actual legally enforced theoretical tech patents for that, and even then they probably wouldn't bite and would deter other actual implementors as you point out. The only path I can see to putting private money into theoretical technology, besides the charity model of Foresight/IMM etc. (not much money there) might be prediction markets as I suggested in my previous comment, to the extent if any those pan out for the purpose of betting on future uncertainties.

12:24 AM

A said...

Ah, I did misread. I'm sorry about that -- I tend to think of patent law in terms of reform proposals, and that was obviously in the front of my mind.

Putting aside the other issues, I wanted to clarify the VC/incubator notion for a moment. This might be a separate track, but it's also related to what you're discussing.

As you say, most VC firms and most incubators are almost solely concerned with operationalizing and implementing technologies, not with R&D. That's the primary function of many incubators, especially those affiliated with universities -- to take ideas that are usually already patentable and link them up with developers, managers, etc.

Of course, an incubator focused on making patentable, specifiable ideas would really be nothing more than an R&D group, although there may be some fertile ground for making more "general-purpose" R&D projects -- I'll have to go look at the literature.

What I was thinking about (and this is the issue that's strongly alluded to in your post) is an incubator as a risk-sharing mechanism between entrenched actors in a market. The whole "get bought up by Google" notion that's been floated by Paul Graham the last couple of years is what I'm thinking of -- it creates a market that's highly attractive to both entrepreneurs and to established companies (who get to stay ahead of the "Innovator's Dilemma" of disruptive technologies).

Of course, there's often projects that come along, like YouTube or MySpace, whose value is highly inflated by the fact that they're not replicable -- a "group fund" for R&D and purchasing might be a good way to ameliorate this imbalance problem.

That is, a group fund between Google, Yahoo!, and whoever else targeted at purchasing those companies would decrease the price (because of fewer competitive offers) and hopefully free up more capital for acquisitions lower down the ladder (an advantage for entrepreneurs in general). And since many of these services aren't being bought for their market share rather than as pure assets for the new companies, it seems like the big companies wouldn't care too much about profit-sharing (i.e. Google and Yahoo! don't care about how much revenue YouTube brings in -- that's not why they made the acquisition).

Maybe a bit outside the ambit of what you're proposing, but that's what I was thinking about -- a group R&D fund as risk-spreading in first-past-the-post sorts of markets. Just wanted to get it off my chest, and apologies again for the misread :)

7:28 PM

nick said...

"...a group fund... targeted at purchasing those companies would decrease the price (because of fewer competitive offers) and hopefully free up more capital for acquisitions lower down the ladder (an advantage for entrepreneurs in general)."

That sounds like a good economic argument, although legally one might have to worry about anti-trust issues.

Even if the acquirers don't care about current revenue they do care greatly about revenue potential, e.g. eyeballs. In these kinds of startups the biggest uncertainties are usually social factors, like how many people one can draw to participate in and help build the community. An even more unpredictable "art" than chemistry (to tie it into my post, as I didn't see how you were doing that :-)

9:53 PM

A said...

Well, I accept your challenge to invent some post ex facto relevance! ;)

What I was getting at is that I think your proposal has merit, but perhaps not on an individual basis. If we're talking about an R&D methodology here, then I think there's still an "Innovator's Dilemma" problem. What you're proposing is a way of quantifying and managing R&D risks, but I'm not sure that's what the problem you're identifying really is.

To make an analogy, your proposal is something like the Li formula in derivative markets -- it allows you to more accurately assess risk, but (unlike in a pure market) an established company usually can't jump out of its current risk tranche into the high-risk tranche where all the disruptive technologies are being developed.

This is especially the case in a market where all the R&D budgets are pitted against each other in a first-past-the-post system (especially true in chemistry, biology, and the other 'predictable' fields -- the first breakthrough creates a patent monopoly).

I think that, as you suggest somewhat obliquely, there's also a problem with the current market adopting predatory techniques toward new technologies (why do we have anti-trust legislation?) -- and better risk-evaluation methods are conducive to box-out strategies.

Now, it seems to me that all three of those problems can be addressed by industry-specific incubators. For example, in the solar-power field, there might be a significant need for a very specific chemistry breakthrough, and if the risk is predictable, it makes sense for the current companies to jointly fund the development and attract the disruptive players to the market in an amenable way. (Like I said, much like the 'Web 2.0' market has been working recently.)

So the specific interplay between what you're talking about and what I'm talking about is that the risk-evaluation formula is critical as an enabling factor, especially since we're talking about loss-prone, high-risk operations like VC funds and incubators. Think of what credit derivatives looked like before Li and the Gaussian copula came along -- relatively primitive. You need both the formula and the model -- I think that's the analog. At any rate, I think that you have to get through the hoop of competing R&D budgets, and accurate risk-management is only the first step.

2:32 PM

Nick Szabo said...

This comment has been removed by the author.

4:10 AM

test said...

test

8:25 PM

test said...

test

1:51 PM

Post a Comment

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.